REMARKS

Claims 1-4, 7-19, and 21-38 are pending in the application. Claims 1, 19, 23, 24, and 25 are currently amended.

The Examiner is thanked for granting the interview on August 8, 2007.

In the Interview the issue of the use of prior knowledge about the structure of the document being analyzed was raised (as described in the disclosure in numerous references to "intelligent structure analysis"). In light of the interview, Applicants now add clarifying language to the independent claims to indicate the use of said prior knowledge during the analyzing.

In addition, the issue of the internal publication format being in a mark-up language (as previously claimed in dependent claim 2, now cancelled), such as XML was discussed.   In light of the interview, Applicants now add the language "said internal publication format comprising a mark-up language" to all independent claims.

Claim Rejections – 35 USC 103

In this section of the Office Action, claims 1-4, 7-19, and 21-38 were rejected under 35 USC 103(a) as being unpatentable over Ferrel et al., US006199082B1 in view of Cooperman US006377704B1.

Favorable reconsideration of this rejection is respectfully requested since, as will be shown below, the above-amended and previously presented claims are both novel and inventive over the prior art cited by the Examiner.

The present disclosure teaches a system that automatically carries out intelligent structure analysis of a scan or like existing passive digital version of an existing document. The scan, or like existing passive digital version, preserves the layout of the document, but has no further structural or other information or metadata to differentiate any portion of the document from any other portion. While the passive digital version provides an image of the page layout for the human eye it has no meaning for computer programs.

Within the passive digital version of an existing document, no object is distinguishable from another – indeed, no object is identifiable as an object per se. Rather, the passive digital version is a collection of undifferentiated bits and bytes.

Not only are objects within the existing passive document not recognized, neither are any relationships that may exist among them since, as noted, the objects themselves are not identifiable as objects.

No semantic nor structural nor organizational information is present to permit any sort of consideration or operation on meaningful subsets of the document (where "meaningful subset" is another name for an "object" or "objects" as described in the application).

Examples of desirable consideration and operations are

- search and retrieval of specific objects in the documents;
- content-sensitive and structural navigation;
- re-organization of the document's component parts;
- re-publishing all or a subset of the document in a different presentation format;
- re-publishing all or a subset of the document on a different platform;
- identification of relationships between subset objects, such as an article and its accompanying illustration, headings etc., or part A of an article with part B of the same article;
- identification of types of objects for additional, object-specific processing, such as, but not limited to, object-specific compression, OCR, and semantic analysis;
- identification of relationships to other documents.

An aspect of the present invention is to infuse meaning and structure to an existing document, where such meaning and structure is lacking prior to the operation of an embodiment of the invention.

The steps to achieve an enhanced document augmented with metadata fully describing the components of the document and the relationships among them (and back to the original passive document) are described in the application. This is no mere transformation of data, but involves sophisticated analysis of the passive document far beyond basic segmentation into rectangles or OCR of textual regions.

The results of the sophisticated intelligent analysis are stored in a repository that includes various transformations of the original passive document (e.g., cropped and at different resolutions) as well as the metadata for each component derived by the analysis. The metadata incorporates simple information, such as references to the

original font and size (for a text object) as well as <u>intricate details of the objects' properties and relationships to other objects</u> as well as <u>relationships to the full source page from which it was extracted</u>

A product of the present invention is the use of the data stored in the repository to republish the original document's contents in new and innovative ways. For example, an embodiment of the system can take a very large collection of archived newspapers scanned from where they were previously stored on microfiche and automatically provide them as a collection of indexed, searchable, stand-alone articles, including their associated graphic illustrations and links between related articles. No such ability is taught in the prior art.

The foregoing is only one example of the benefits of the present invention. Additional benefits are described in the application and others will also be evident to one skilled in the art.

Prior art cited by the Examiner

The Applicants respectfully thank the Examiner for the new citation of the prior art taught by Cooperman US006377704B1 – filed 4-30-1998 (hereinafter Cooperman) and welcome the opportunity to demonstrate how the present invention is both novel and inventive over the teachings of Cooperman and other prior art cited by the Examiner, namely Ferrel, et al US006377704B1 – filed 07-17-1995 (hereinafter Ferrel).

Examiner has rejected all claims as being unpatentable due to obviousness by Ferrel in light of Cooperman. The applicants respectfully traverse and submit that Examiner has not made out a *prima facie* case of obviousness as demonstrated in the detailed discussions below.

Additionally, Examiner has stated "Using the broadest reasonable interpretation, the Examiner equates the <u>claimed</u> **converting data from one format to a different format** as equivalent to a set of predefined styles along with supporting macros."

Applicants respectfully refer to claim 1 wherein is claimed **converting each object to an internal publication format** and to claim 19 wherein is claimed a **mark-up language distiller module for converting the newspaper from said original format to a mark-up language format and a publisher server for**

12

converting the data from said mark-up language format to a final publication format, and to claim 23 wherein is claimed converting each object to an internal publication format and to claim 24 wherein is claimed converting each object to an internal publication format and to claim 25 wherein is claimed converting each object to an internal publication format.

As noted, Applicants have not claimed converting data from one format to a different format and said language does not appear in Applicants' claims. Applicants have claimed only the limited cases shown above drawn from the independent claims (along with additional limitations in the relevant dependent claims thereto). Neither Ferrel nor Cooperman teaches nor even hints at the idea of converting data *from* the various sources enumerated in the claims *to* an internal publication format or a mark-up language format.

In light of the Examiner's apparent misreading of Applicants' claims, Applicants respectfully request withdrawal of the finality of the action.

Ferrel discloses a multimedia publishing system for creating online publications, wherein the contents and design aspects of the online publishing process are separated into distinct components before being transmitted to the customers' computers. Separation of the contents and the design aspects of the online publications facilitates efficient authoring of documents by independent content-creators and design-creators as well as efficient distribution of published documents by enabling the transmission of, for example, high-quality titles over low-speed communications links subject to loss of connectivity.

Cooperman teaches an innovation in the realm of OCR, specifically, identifying insets in a scanned document so as to facilitate determining correct reading order and creating, for example, an editable approximation of the original document.

As will be shown in the detailed discussions below, the present invention is both novel and inventive over the teachings of Ferrel and Cooperman, as well as the combination thereof, should such a combination be considered likely by one trained in the art of either of the fields to which said teachings relate.

The Examiner has suggested that one of ordinary skill in the art would have combined the teachings of Ferrel with those of Cooperman. The Applicants respectfully traverse. Ferrel teaches a system for publishing data specifically for authoring new, original content where design elements are authored and managed distinctly from content: "One unique concept that permeates the MP system is the clean separation of content and design. In this context, content is defined as the actual data that is to be displayed to the user. The design of a project is how that information gets displayed to the user (e.g., its format on the computer screen)." (Ferrel col 8, lines 4-9). This is reinforced by Ferrel's observation that "...the MP system supports standards such as Microsoft Word and Standard Generalized Markup Language (SGML) to ensure that the content provider's investment in existing tools can be fully leveraged. The MP system also reads standard HyperText Markup Language (HTML) documents so that existing HTML documents can be easily converted to more sophisticated applications. Additionally, through support of the OLE standard, tools that supports OLE server capabilities can be used to create content embedded in an MPS title. By supporting additional standard file formats, the MPS can also accommodate other tools (for example high-end graphic applications)." (Ferrel col. 62, line 66 to col. 63 line 10), emphasis added). Ferrel's "existing tools" are ones for creating new content. The Ferrel system "reads", i.e., imports HTML and can accommodate "standard file formats". None of these suggest the use of intelligent structure analysis on pre-existing documents.

Cooperman, by contrast, teaches a minor improvement to OCR systems, namely inset detection to improve determination of correct reading order and producing editable documents in correct reading order (see above). The teachings of Cooperman are not related to the field of modern publishing systems, but rather to the conversion of pre-existing documents to machine-readable form, e.g., for archival purposes.

One trained in the art of archival analysis is not typically also involved in the field of modern publishing systems and vice versa. Thus, the teachings of Cooperman and Ferrel would not ordinarily be combined.

In any case, as shown above and further shown in the detailed discussion of the claims below, the teachings of both Ferrel and Cooperman lack inventive and innovative teachings of the present application – said inventive and innovative teachings are lacking from the combination as well.

As one specific example, both Ferrel, and Cooperman, and the combination thereof, lack automatically analyzing the newspaper to decompose the predetermined layout of each page of the newspaper in the original, existing format into said plurality of blocks, each block representing an object; said analyzing furthermore identifying at least one logical relationship between said blocks; converting each object to an internal publication format, said internal publication format identifying and preserving said internal structure of said blocks within said objects, said internal publication format furthermore preserving said layout as a relationship between said objects; said internal publication format furthermore preserving said logical relationship; and rendering said internal publication format to incorporate said objects, said layout and respective internal structures and said logical relationship in the final publication format.

The aforementioned lack in the cited prior art, as well as additional new and innovative teachings of the present invention are described in detail in the sections following.

Claim 1, as currently amended, defines a method for automatically publishing data in a final publication format, wherein the data is in the form of a newspaper having an original, existing format including a plurality of pages, each page having a predetermined layout comprising a plurality of independently standing data blocks, each block having an internal structure, the method comprising: obtaining a scanned representation of said newspaper, said representation preserving said layout, from said representation automatically analyzing the newspaper to decompose the predetermined layout of each page of the newspaper in the original, existing format into said plurality of blocks, each block representing an object; said analyzing furthermore applying prior knowledge of newspaper structure and identifying at least one logical relationship between said blocks; converting each object to an internal publication format, said internal publication format identifying and preserving said internal structure of said blocks within said objects, said internal publication format furthermore preserving said layout as a relationship between said objects; said internal publication format furthermore preserving said logical relationship; and said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure and rendering said internal publication format to

15

incorporate said objects, said layout and respective internal structures and said logical relationship in the final publication format.

Regarding **independent claim 1**, Examiner has stated (pg. 2) that "Ferrel teaches from said representation automatically analyzing the newspaper to decompose the predetermined layout of each page of the newspaper in the original, existing format into said, plurality of blocks, each block representing an object; converting each object to an internal publication format, said internal publication format ide ntifying and preserving said internal structure of said blocks within said objects, said internal publication format furthermore preserving said layout as a relationship between said objects. ... Using the broadest reasonable interpretation, the Examiner equates the claimed **internal publication format** as equivalent to a set of predefined styles along with supporting macros."

Applicants respectfully traverse. Ferrel teaches a system for <u>authoring</u> and publishing data. The Ferrel system separates the design and layout definitions from the content definition: "One unique concept that permeates the MP system is the clean separation of content and design. In this context, content is defined as the actual data that is to be displayed to the user. The design of a project is how that information gets displayed to the user (e.g., its format on the computer screen)." (Ferrel col 8, lines 4-9).

This separation facilitates <u>authoring by authors who have expertise in layout design</u> independently of <u>authoring by authors who have expertise in the content</u> (Ferrel col. 63, lines 13-16). The separation also facilitates publication where style and layout information may be communicated to a remote device one time, but used many times for varying and updated content, thus <u>optimizing use of low-speed communication links</u> (Ferrel col. 62, lines 53-65).

Ferrel does not teach, nor even hint at, from said representation automatically <u>analyzing</u> the newspaper to <u>decompose</u> the predetermined layout of each page of the newspaper in the original, existing format into said plurality of blocks, each block representing an object. <u>The Ferrel publishing system relies upon the authors to manually define elements and indicate relationships</u>, if any, between elements, whereas the present invention performs intelligent and <u>automatic</u>

16

structure analysis to <u>discover</u> elements and relationships between elements. (For the discussion related to the teachings of Cooperman, please see below.)

Moreover, Ferrel does not teach, nor even hint at, <u>converting</u> **each object to an internal publication format**. Since Ferrel does not teach **decomposing** an original document in the first place, Ferrel certainly does not teach any further operations thereupon. I.E., **objects** of this claim do not appear in Ferrel's teachings, so a conversion of those objects certainly does not appear.

In light of the Examiner's comments in the interview, Applicants have added clarifying language to the claim regarding applying <u>prior knowledge of newspaper structure</u> and **identifying at least one logical relationship between said blocks** and **said internal publication format furthermore preserving said logical relationship and rendering said internal publication format to incorporate ... said logical relationship**. This is described in the present application in several places (e.g., "XML distiller module 18 preferably performs <u>intelligent</u> structure analysis, in order to be able to recognize and define the structures and objects contained in the newspaper data" [0036]; "Recognition of the <u>hierarchy</u> for newspaper objects is an important part of the process" [0037]; and "the present invention does not merely convert the data from one format to a different format, but instead is able to also provide at least <u>a basic structure for organizing</u> the data" [0024].

Ferrel does not teach, nor even hint at said <u>**internal publication format identifying and preserving**</u> said internal structure of said blocks within said objects, said internal publication format furthermore <u>preserving</u> said layout as a relationship between said objects; said internal publication format furthermore preserving said logical relationship. The structure and relationship identifying and preserving here claimed relate to said objects derived in the analyzing. Ferrel does not teach, nor even hint at, said **analyzing** and certainly not **identifying and preserving**.

In addition, in light of the Examiner's comments during the interview, Applicant has incorporated the language of dependent claim 2 (now canceled) to limit claim 1 to a method with said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure . As a further clarification, Applicant respectfully points out that XML (as claimed in dependent claim 3) is but one example of mark-up languages suitable for the purpose. Other markup languages currently available that may be appropriate include HTML,

SGML, XSL, and others. It is understood that one trained in the art may also use other markup languages including ones specifically created for the purpose of the method described in claim 1. This is discussed in Applicants' original disclosure in numerous places including, inter alia, "Repository 20 preferably stores these information objects in two different data formats: a suitable Web page and/or document mark-up language format, such as XML for example" [0045]; "A document mark-up language such as XML is particularly preferred for storing this type of information about the entity," [0047]; "As a non-limiting example, the operation of the present invention with regard to z-order may be optionally performed by using a mark-up language such as HTML for example." [077];

Applicants respectfully submit, therefore, that the claimed **internal publication format** is distinct from the teachings of Ferrel, whose "set of predefined styles along with supporting macros" (Ferrel col. 26, lines 42-43) do not include the claimed **relationships, blocks** and **objects**, nor an **internal publication format** comprising a mark-up language.

In the Ferrel system, all content data is <u>newly created</u> by authors exercising creativity to create new content. In the presently discussed claim 1 is claimed **converting** each object to an internal publication format. Such **converting** is not taught nor hinted at by Ferrel since in the Ferrel system the data is created by authors in a publication format from the outset. As Ferrel states: "The MPS authoring environment provides a way for <u>authors to create structured documents</u>" and "To create content for the MP system 100 in the MPS Document Editor 188, <u>an author creates a document</u> based on the MPS template." (Ferrel col 16, lines 19-42) Furthermore, Ferrel teaches conversion from one <u>publication</u> format (e.g., Microsoft Word, SGML, HTML, OLE (Ferrel col 62 line 66 to col 63 line 10)) to another publication format, but *not* **converting each object** derived from <u>said</u> representation to an internal publication format.

Examiner has also stated (pg. 4) that "Ferrel teaches **rendering said internal publication format to incorporate said objects said layout and respective internal structures in the final publication format**" citing Ferrel's teachings about style sheets, styles, markup languages, and various output devices.

Applicants respectfully traverse. As shown above, the claimed **internal publication format** is distinct from the teachings of Ferrel and the claimed **rendering** therefrom must also be distinct.

Examiner has further stated (pg. 4) that "Ferrel does not expressly teach, but Cooperman teaches **obtaining a scanned representation of said newspaper, said representation preserving said layout.**" Applicants respectfully traverse.

Cooperman provides a cogent statement of a problem faced in the art of OCR (col. 1, lines 43-49):

> Document layout analysis is a process by which the information regarding the organization of the document content, i.e. its structure, is extracted from the document image. The structure identifies document entity types (e.g. paragraphs, figures and tables), their properties (e.g. the number of columns in a table), and their interrelations (e.g., a figure is above a caption).

Cooperman continues and explains that the art of OCR is concerned with this problem specifically to solve the problem of producing a document in correct reading order. As Cooperman states (col. 1, lines 57-61) (emphasis added):

> Indeed, such information is important to state-of-the-art OCR systems that attempt to produce textual output in "reading order" (e.g., appropriately reflecting the flow of text in columns, headers, footers, captions and now insets).

As Cooperman makes clear, the problem addressed by Cooperman's disclosure and invention is specifically in the realm of producing a document in "reading order", and, more specifically, to addressing the problem of insets.

Cooperman later explains that the general problem of document layout analysis is not solved in the art (col. 6, lines 14-26) (emphasis added):

> For each region or segment identified within a particular page of the document, the following information is preferably provided by a segmentation system (not shown):
>
> 1) A description of each region as either text, graphic (e.g., picture, drawing, etc.), ruling, timestamp, or table cell;
>
> 2) The coordinates of the region's bounding box in absolute distance from either the scanned edge or the hard copy document's edge; and

19

3) For text regions, font information comprising: (a) the most common font, (b) the font size, (c) the font style (e.g., regular, bold, italic, fixed pitch, etc.), and (d) the number of lines of text within the region.

As Cooperman states, the low-level identification of regions as text, graphic, etc., the coordinates thereof, and partial ancillary information related to the fonts <u>does not constitute "document layout analysis" per se, but rather "segmentation."</u> Indeed, Cooperman's disclosure continues by disclosing a contribution to solving a small part of the <u>document layout analysis problem</u>, when stating (col. 6 lines 27-36) (emphasis added):

In response to the above inputs and in accordance with the present invention, a <u>document layout analysis</u> system <u>with inset detection capability</u> preferably provides the following outputs:

1) An identification of insets within the document;

2) A classification of the inset type (e.g., frame, credit, center, column or stray; examples of frame, center and column depicted in FIGS. *6-8)*;

3) The reading order of text regions;

4) A grouping of text regions into sections.

In addition to clarifying the distinction between mere <u>segmentation</u> and <u>document layout analysis</u>, Cooperman's disclosure makes clear that the contribution to the art therein is one of identifying insets and establishing correct reading order. As stated, this is "...to produce the output of the document layout analysis system-- information sufficient to recreate a close approximation of an original document <u>so that the information may be employed to produce a formatted file (e.g., word processor format 76c in FIG. 2) that is editable</u> and bears a close approximation to the appearance of the original document." (col. 6, lines 44-50) (emphasis added)

Furthermore, it is clear that when Cooperman states that a product of the process is a file that "bears a close approximation to the appearance of the original document" there is <u>no intent to completely preserve the original document's layout</u>. Indeed, both <u>the inputs and outputs are insufficient to do so</u>. The only intent is to achieve an "<u>approximation</u>" (e.g., "the most common font" and not all font information), and not an exact replica with added features.

Significantly, the stated goal of Cooperman's embodiments is "to produce a formatted file...that is editable." The Cooperman disclosure does not teach nor even hint at creating a fully decomposed and annotated (i.e. marked-up) intermediate representation suitable for re-publication with interactive features in both the original form and arbitrary and multiple new forms, as claimed in the currently amended claims. Cooperman also does not teach nor even hint at re-publication with interactive features as enabled by said analyzing furthermore applying prior knowledge of newspaper structure and identifying at least one logical relationship between said blocks and internal publication format furthermore preserving said logical relationship and said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure in either the original form or arbitrary and multiple new forms, as claimed in the currently amended claims.

As described in the application, embodiments of the present invention provide re-publication of the original document with interactive features in both the exact original form and arbitrary and multiple new forms based upon a fully decomposed and annotated (i.e. marked-up) intermediate representation. Furthermore, the intelligent layout analysis employed by embodiments of the present invention results in identification of high-level logical objects in the original document (e.g., an "article, an advertisement, a picture which is not otherwise associated with an article or advertisement, and general data, which covers information that does not fit into one of the preceding categories. Examples of general data include, but are not limited to, crossword puzzles, television guide listings, stock quotes and so forth." [0036]. Other types of identified and classified objects mentioned (but not limited to these) in the application include: Titles, subtitles, Roof title, Abstract, Bylines, Lead, Continuation, Captions, Graphic elements (lines, frames, decorations), Sections, Subsections.

The intelligent layout analysis also goes further in analyzing objects and decomposes them into primitives, which are also stored in the repository ("A primitive is a rectangular region of a printed newspaper page, which contains textual and/or graphical information. Articles and other entities on the printed newspaper page are optionally and preferably represented as a "puzzle", or structured collection, of the rectangular primitives from which the printed page is composed." [0047] "Primitives preferably comprise lines, words, and quads (which together form broken or multiple-style words). [0048]"

Another feature of the present invention is that it analyzes the elements so as to identify each element and how each element is logically related to all other elements. For example, it may be determined that a certain text belongs to a certain picture. A new interactive version of the document, or subset of the document, can be published that allows one to click on the picture and get the related text. Cooperman does not teach, nor even hint at, **automatically analyzing the newspaper to decompose the predetermined layout of each page of the newspaper in the original, existing format into said plurality of blocks, each block representing an object; said analyzing furthermore applying prior knowledge of newspaper structure and identifying at least one logical relationship between said blocks; converting each object to an internal publication format, said internal publication format identifying and preserving said internal structure of said blocks within said objects, said internal publication format furthermore preserving said layout as a relationship between said objects; said internal publication format furthermore preserving said logical relationship; and said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure and rendering said internal publication format to incorporate said objects, said layout and respective internal structures and said logical relationship in the final publication format.**

This is discussed in the present application, specifically at [0025], wherein it is stated that "the structure of the published content uses the interactive capabilities of the Web page to their fullest extent, by enabling the user to view different portions of the content in any desired order, with links between these portions being made according to the information in the content and/or according to the type of content, rather than according to an arbitrary structure imposed according to the requirements of the printed newspaper" and at [0029] "This process results in a set of enhanced, structured files which combine the original image of the data, optionally in the TIF format as previously described, but preferably in the GIF and/or JPG formats, with the text and XML information. Each such file thus preferably maintains the visual aspects of the newspaper layout, while enabling far greater functionality to be available through the Web page version of the newspaper" and [0036] "Next, XML distiller module 18 preferably performs intelligent structure analysis, in order to be able to recognize and define the structures and objects contained in the newspaper data,

particularly with regard to each page of the newspaper" and [0046] wherein is stated that for the entities stored in the repository "the mark-up language format, such as XML, preferably contains the actual content of the newspaper object, as well as details of the internal structure and relationship(s) between its various components," (emphasis added)

The automatic recognition of relationships also includes structural relations, such as that stated in [0037]: "Recognition of the hierarchy for newspaper objects is an important part of the process. One illustrative, non-limiting example of such a hierarchy for recognized newspaper objects is the following: Newspaper→Section→Page→Subsection→Article→Photograph→Caption→Photographer (by name)."

In summary, as clearly illustrated by the above examples, embodiments of the present invention perform, inter alia, intelligent structure analysis providing the identification of high-level objects and primitives in the original document as well as numerous and complex relationships between the objects, which is not taught in the cited prior art.

With regard to Cooperman specifically, Cooperman does not teach, nor even hint at, said analyzing furthermore identifying logical structures and relationships between blocks and objects as claimed in the currently amended claim.

Thus, even if the teachings of Ferrel were combined with those of Cooperman, one would not obtain the method of the current claim.

It is thus believed that claim 1 is both novel and inventive over the prior art and respectfully maintained that the claim should be allowed.

Claim 19, as currently amended, defines a system for automatically publishing received data of a pre-existing newspaper, the newspaper in a computerized format, the system comprising: (a) at least one source of newspaper data, said source preserving an original, structure of the newspaper, the computerized format comprising a plurality of pages, each page having a predetermined layout comprising a plurality of independently standing data blocks, each block having an internal structure; (b) a mark-up language distiller module for converting the newspaper from said original format to a mark-up language format, wherein said mark-up language distiller module applies prior knowledge of newspaper

structure and automatically analyzes the newspaper data in said original, existing digital format to (1) decompose the newspaper data into said plurality of blocks, each block with said internal structure representing an independent data object, each object having content and at least one attribute of the data, such that each object is converted to said mark-up language format, said markup language distiller module further analyzing and preserving said structure; and (2) identify at least one logical relationship between blocks, said markup language distiller module further analyzing and preserving said logical relationship; and (c) a publisher server for converting the data from said mark-up language format to a final publication format, said final publication format incorporating, as objects, said blocks with said internal structure, and further defining interrelationships with said objects.

Regarding independent claim 19, Examiner has stated (pg. 5) that "Ferrel teaches a mark-up language distiller module for converting the newspaper from said original format to a mark-up language format, wherein said mark-up language distiller module automatically analyzes the newspaper data in said original, existing digital format to decompose the newspaper data into said plurality of blocks, each block with said internal structure representing an independent data object, each object having content and at least one attribute of the data, such that each object is converted to said markup language format, said markup language format further analyzing and preserving said structure", and (pg. 6) "Using the broadest reasonable interpretation, the Examiner equates the claimed a markup language distiller and internal structure representing an independent data object as equivalent to a set of predefined styles along with supporting macros."

Applicants respectfully traverse. As discussed above with regard to claim 1, Ferrel's system is one where authors create new content. Ferrel does not teach, nor even hint at, a system for automatically publishing <u>received</u> data of a <u>pre-existing</u> newspaper, nor does Ferrel teach, nor even hint at, <u>any</u> mechanism that <u>automatically analyzes</u> the newspaper data in said original, existing digital format <u>to decompose</u> the newspaper data into said plurality of blocks, each block with said internal structure representing an independent data object, each object having content and at least one attribute of the data.

In light of the Examiner's comments, Applicants have added clarifying language to the claim regarding how a mark-up language distiller module applies

prior knowledge of newspaper structure and automatically analyzes the newspaper data in said original, existing digital format to ... identify logical relationships between blocks and objects, said markup language distiller module further analyzing and preserving said logical relationships. This is described in the present application in several places (e.g., "XML distiller module 18 preferably performs intelligent structure analysis, in order to be able to recognize and define the structures and objects contained in the newspaper data" [0036]; "Recognition of the hierarchy for newspaper objects is an important part of the process" [0037]; and "the present invention does not merely convert the data from one format to a different format, but instead is able to also provide at least a basic structure for organizing the data" [0024].

Applicants have also taken advantage of this opportunity to edit the earlier language in the claim to clarify that it is the said markup language distiller module further analyzing and preserving said structure.

Ferrel's "set of predefined styles along with supporting macros" are provided by a "template" used by an author to create a new document (Ferrel col 26, lines 42-43). The claimed internal structure is used to represent[ing] an independent data object derived as a result of the process where said mark-up language distiller module applies prior knowledge of newspaper structure and automatically analyzes the newspaper data in said original, existing digital format to decompose the newspaper data into said plurality of blocks, each block with said internal structure representing an independent data object, each object having content and at least one attribute of the data, such that each object is converted to said markup language format, said markup language distiller module further analyzing and preserving said structure". Ferrel does not teach, nor even hint at, a language distiller module that has prior knowledge of newspaper structure and automatically analyzes a pre-existing newspaper to decompose it into blocks and structures and identify at least one logical relationship between blocks.

The claimed internal structure preserves the representation and structures and relationships derived from the analysis. No such capability is taught by Ferrel, nor, as explained in detail above in the discussion of claim 1, by Cooperman, and certainly not in the combination.

It is thus believed that claim 19 is both novel and inventive over the prior art and respectfully maintained that the claim should be allowed.

Claim 23, as currently amended, defines a method for automatically publishing data of a preexisting document in a final publication format, wherein the data is received in an original, existing format comprising a plurality of pages, each page having a predetermined layout comprising a plurality of independently standing data blocks, each block having an internal structure the method comprising: obtaining a representation of the preexisting document, the representation preserving the layout; from the representation, automatically analyzing the preexisting document to decompose the data received in the original, existing format into a plurality of objects, each object corresponding to one of the blocks; said analyzing furthermore applying prior knowledge of document structure and identifying at least one logical relationship between said blocks; preparing a list of text and/or graphic elements for each object; determining properties of each element, including determining visibility and overlap characteristics for each graphic element within the object; recognizing structural layout properties of the data in an original format; converting each object to an internal publication format; said internal publication format furthermore preserving said logical relationship; and said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure and rendering the internal publication format in the final publication format, the final publication format presenting the blocks as the independently standing objects incorporating the internal structure and the layout and said logical relationship.

As explained in detail above with regard to claim 1, neither Ferrel nor Cooperman teaches the intelligent structure analysis required for **automatically analyzing the preexisting document to decompose the data received in the original, existing format into a plurality of objects, each object corresponding to one of said blocks said analyzing furthermore applying prior knowledge of document structure and identifying at least one logical relationship between said blocks**; and it is certainly not taught in the combination.

Examiner has stated (pg. 9) that "Ferrel teaches **preparing a list of text and/or graphic elements for each object**. Specifically, Ferrel discloses the method of determining properties of each element, including determining visibility (as taught by Ferrell at col. 26, lines 15-25), and overlap characteristics for each graphic element

within said object (as taught by Ferrell at col. 36, line 25-67, and at col. 38, lines 30-45)."

Ferrel's teachings relate to an "Authoring Subsystem" (Ferrel, col. 26, line 13). An authoring system is one used by human authors to create new content. As Ferrel continues (col. 26, line 14-25), "Content is separated from design in the MP system 100. In the Viewer 202 (FIG. 2), content and design are brought together by controls to display a title as specified by the designer. As a result, these controls need to identify different elements in the structure of the content so they may format it correctly. This is done by creating structured content. The MPS authoring environment provides a way for authors to create structured documents. The MPS authoring environment includes the MPS Document Editor 188, which supports the creation of structured documents, insertion of links and the application of properties to these documents for content retrieval."

While Ferrel's described "authoring environment provides a way for authors to create structured documents", claim 23 is for a method for automatically publishing data of a preexisting document in a final publication format, wherein the data is received in an original, existing format comprising a plurality of pages, each page having a predetermined layout. Ferrel does not teach, nor even hint at automatically publishing data of a preexisting document.

Furthermore, Ferrel's teaching "these controls need to identify different elements in the structure of the content so they may format it correctly" has to do with the identification of individual elements as being of one type or another so that "they may format it correctly" and does not teach nor even hint at compiling or in any way preparing a list of text and/or graphic elements for each object.

It is thus believed that claim 23 is both novel and inventive over the prior art and respectfully maintained that the claim should be allowed.

Claim 24, as currently amended, defines a method for automatically publishing received data in a final publication format, the data having an original, existing format comprising a plurality of pages, each page having a predetermined layout comprising a plurality of independently standing data blocks, the method comprising: obtaining a representation of the received data having an original existing format, the representation preserving the layout; from the representation, automatically analyzing the pages to decompose the pages of the data received in the

original, existing format into a plurality of objects, the objects corresponding to the blocks; said analyzing furthermore applying prior knowledge of page structure and identifying at least one logical relationship between said blocks; preparing a list of text and/or graphic elements for each object; determining properties of each element, including determining a special characteristic for each text element; recognizing structural layout properties of the data in an original format; converting each object to an internal publication format; said internal publication format identifying and preserving said internal structure of said blocks within said objects, said internal publication format furthermore preserving said layout as a relationship between said objects; said internal publication format furthermore preserving said logical relationship; and said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure and rendering the internal publication format in the final publication format such as to include the recognized structure in the objects and said logical relationship and to include the layout.

As explained above in detail with regard to claim 1, neither Ferrel nor Cooperman teaches the _intelligent structure analysis_ required for **automatically analyzing** the pages **to decompose** the pages of the data received in the original, existing format into a plurality of objects, said objects corresponding to said blocks; said analyzing furthermore applying prior knowledge of page structure and identifying at least one logical relationship between said blocks; preparing a list of text and/or graphic elements for each object; **determining properties of each element,** including determining a special characteristic for each text element; **recognizing structural layout properties** of the data in an original format nor the use of an internal publication format comprising a mark-up language and these are certainly not taught in the combination.

Other details of the Examiners objections have been addressed above as well.

It is thus believed that claim 24 is both novel and inventive over the prior art and respectfully maintained that the claim should be allowed.

Claim 25 defines a method for automatically publishing received data in a final publication format, wherein the data is in the form of a newspaper, the newspaper having an original, existing format comprising a plurality of pages, each page having a predetermined layout comprising a plurality of independently standing data blocks, each block having structural layout properties, the method comprising:

obtaining a representation of said newspaper, said representation preserving said layout; from said representation, automatically analyzing the newspaper to decompose the data received in the original, existing format into a plurality of objects, said objects corresponding to said independently standing blocks; and said analyzing furthermore applying prior knowledge of newspaper structure and identifying at least one logical relationship between said blocks; preparing a list of text and/or graphic elements for each object; determining properties of each element; recognizing said structural layout properties of the data in an original format; determining each text segment for each object; building a text block from a plurality of aligned text segments; converting each object to an internal publication format; said internal publication format furthermore preserving said logical relationship; and said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure and rendering said internal publication format in the final publication format to comprise said blocks as objects incorporating said structural layout properties, and to include said layout and said logical relationship.

As explained above in detail with regard to claim 1, neither Ferrel nor Cooperman teaches the intelligent structure analysis applying prior knowledge of newspaper structure required for automatically analyzing the newspaper to decompose the data received in the original, existing format into a plurality of objects, said objects corresponding to said independently standing blocks; and said analyzing furthermore applying prior knowledge of newspaper structure and identifying at least one logical relationship between said blocks; preparing a list of text and/or graphic elements for each object; determining properties of each element; recognizing said structural layout properties of the data in an original format; determining each text segment for each object; building a text block from a plurality of aligned text segments; converting each object to an internal publication format nor the use of an internal publication format furthermore preserving said logical relationship; and said internal publication format comprising a mark-up language to indicate said objects and features of said internal structure and these are certainly not taught in the combination.

Other details of the Examiners objections have been addressed above as well.

It is thus believed that claim 25 is both novel and inventive over the prior art and respectfully maintained that the claim should be allowed.

The remaining claims mentioned in this section of the Office Action are believed to be allowable as being dependent on an allowable main claim.

All of the matters raised by the Examiner have been dealt with and are believed to have been overcome.

In view of the foregoing, it is respectfully submitted that all the claims now pending in the application are allowable.

An early Notice of Allowance is therefore respectfully requested.

Respectfully submitted,

Martin D. Moynihan
Registration No. 40,338

Date: August 21, 2007

Encl.:
Petition for Extension (3 Months)
Request for Continued Examination (RCE)